


[Print](#)

Lesson 10: Probability Distributions - Study Notes

Slide 1:

Random Variables

Outcomes of probability experiments that can take on any numerical value are called **random variables**. For example, if we toss a coin 5 times, the number of heads we observe (0 to 5) is a random variable. The average speed (in km/h) of a Formula One race car during a given lap at the Canadian Grand Prix is also an example of a random variable. Numerical random variables can be subdivided into two categories:

- **Discrete Random Variable:** a quantitative random variable that can assume a limited (countable) number of values; e.g. number of heads when flipping a coin 5 times.
- **Continuous Random Variable:** a quantitative random variable that can assume an uncountable number of values; e.g. average speed of car during a given lap.

Slide 2:

Discrete Random Variable Probability Distributions

When simultaneously tossing 2 coins there are 3 possible outcomes when looking for "heads": 0, 1, or 2. Since the possibilities are TT, HT, TH or HH, the probabilities are:

align="center" cellpadding="0" cellspacing="0" bordercolor="#063763" id="tableData">

P (X = 0)	1 / 4
P (X = 1)	1 / 2
P (X = 2)	1 / 4

When these values are presented in a table format, we refer to it as a probability distribution. The values of the distribution are **mutually exclusive (unique)** and exhaustive (describe all possible outcomes) events. The sum of all the possibilities is always 1.

align="center" cellpadding="0" cellspacing="0" bordercolor="#063763" id="tableData">

X	P(X)
0	0.25

1	0.50
2	0.25
Total	1.00

Slide 3:**Mean and Variance**

It is possible to describe the mean and standard deviation of a discrete probability distribution. The mean of a discrete random variable is found much like the mean of a frequency distribution.

$$\mu = \frac{\sum Xf}{n} = \sum (X \times f/n)$$

Since f/n is $= P(X)$, then:

$$\mu = \sum [x P(X)]$$

The mean of a probability distribution is calculated using the same formula we used for the calculation of the mean of a frequency distribution, except that we have replaced the frequency with probability. The mean of X , when dealing with random variables, is known as the Expected Value ($E(X)$), since it represents what we expect to occur in the long run. The variance ($V(X)$) can be calculated using this formula:

$$\sigma^2 = \sum [x^2 P(x)] - \mu^2$$

Slide 4:**Mean and Variance (Cont'd)**

The easiest way to calculate these values is to arrange the data into specialised tables. For example, let's say we have a company which manufactures the testosterone-enhancing drug known as "androstenedione".

With extensive market research, you have assigned the following probabilities for obtaining professional sport team contracts for your product:

align="center" cellpadding="0" cellspacing="0" bordercolor="#063763" id="tableData">

Number of Contracts	style="TEXT-ALIGN: center">Probability
--------------------------------	--

5	0.2
10	0.1
15	0.4
20	0.2
25	0.1
Total	1.0

Now, although we could take this data and do some number crunching, it would be much easier to set up a table which will help us incorporate the formulas for the calculation of the mean and variance.

align="center" cellpadding="0" cellspacing="0" bordercolor="#063763" id="tableData">

x	P(x)	x P(x)	x^2	$x^2 P(x)$
5	0.2	1	25	5
10	0.1	1	100	10
15	0.4	6	225	90
20	0.2	4	400	80
25	0.1	2.5	625	62.5
Total (Σ)	1	14.5	1375	247.5

(This is an example of a complete probability distribution table)

The mean of the distribution can be found as the total for the "x P(x)" column, 14.5. The variance, according to our newest equation, can be calculated by subtracting μ^2 from the value found in the last column.

The variance in this problem is $247.5 - (14.5)^2 = 37.25$. The standard deviation(s) = 6.10.

What this means is that, on average, we will sell 14.5 contracts. This will aid us in determining profit margins, projected sales, etc.

Slide 5:

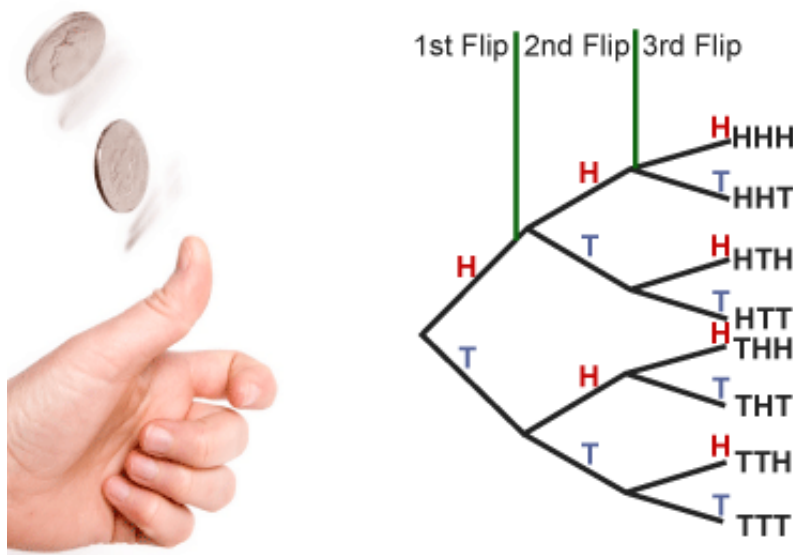
The Binomial Probability Distribution (Discrete)

A Binomial Probability Experiment, also known as a **Bernoulli Trial**, is an experiment made up of repeated trials that have only two possible outcomes (pass/fail, right/wrong, 0 or 1, etc.). In order to be classified as a Binomial Probability Experiment, the test must satisfy the following criteria:

1. There are only two possible outcomes (success/failure).
2. There are n repeated independent trials.
3. The binomial random variable " x " represents the number of successful trials and can take on any integer from 0 to n .
4. Probability of a success = p ; Probability of a failure = q ; where $p + q = 1$ or $p = 1 - q$.
The probability of a success remains the same in all trials.

Each trial is an 'experiment' with exactly 2 possible outcomes, "success" and "failure" with probabilities p and $1-p$, respectively.

Consider a fair coin where the probability of a "head" is 50%. If we flipped this coin 3 times, what are the various probabilities associated with each outcome (0 heads, 1 head, 2 heads, 3 heads)? The easiest way to complete this problem would be via a tree diagram:



By flipping the coin 3 times, there are a total of 8 outcomes, each with their own probability of occurring.

However, what if we flipped the same coin 20 times?

The tree diagram produced here would be quite the project.

Slide 6:

The Binomial Probability Distribution (Discrete) (Cont'd)

Luckily, as we saw earlier in this section, there exists a counting method to alleviate our pain. The combination formula is employed as a tool to calculate the amount of possible combinations given specific criteria. Known as the binomial coefficient, this formula will calculate the number of

possible combinations with a specific number of "successes". In our case, the exact number of "heads" when a coin is flipped n times.

Recall the formula for combinations:

$$C_n^p = \frac{n!}{p!(n-p)!}$$

The **binomial coefficient** uses the same formula but the notation is a little different. Instead of using the "p" (which stands for permutations), the coefficient employs an "x". Furthermore, the C (combinations) is dropped from the expression entirely. Long story short, the binomial coefficient is expressed this way:

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}$$

Where the "n" represents the number of trials, and "x" is the number of successes.



Video Examples

If the probability that a newborn baby will be a girl is 60%, what is the probability that in a family of 3 children:

- All children are the same gender?
- Exactly 2 are girls?
- At least one child is a boy?

If the probability that a newborn baby will be a girl is 60%, what is the probability that in a family of 10 children:

- Exactly 7 are boys?

If the probability that a newborn baby will be a girl is 60%, what is the probability that in a family of 6 children:

- At least 4 are girls?

Hint: The formula needed to solve this problem most efficiently is presented on the next slide.

Solutions

This video provides additional explanations about the binomial distribution function. (Length 17:14)

- View [streaming video](#) if you have a high-speed Internet connection.
- [Download video](#) (311 Mb) if you have a low-speed Internet connection.

Slide 7:

The Binomial Probability Distribution (Discrete) (Cont'd)

Example 1

Let's give this formula a try: If we flipped a coin 20 times (assuming that it is fair), how many combinations will yield **exactly** 15 heads?

$$\binom{20}{15} = \frac{20!}{15!(20-15)!} = 15504$$

There are exactly 15,504 combinations in which there are exactly 15 heads if we flip a coin 20 times! But what if we wanted to know the amount of combinations that would yield **at least** 15 heads? Since this formula gives us the combinations for exactly 15, we would have to repeat it for $x = 16, 17, 18, 19$, and 20. Then we would add up the combinations. If we actually were to do this, we would find 21,700 possibilities (see below for work).

The probability for each combination (exactly 15 out of 20) is the same, and can be calculated by the multiplication rule.

$$\left(\frac{1}{2}\right)^{15} * \left(\frac{1}{2}\right)^5 = 0.000000953$$

P(15 heads) x P(5 tails)

This formula is known as the **binomial probability function** and can be described as:

$$P(x) = \binom{n}{x} (p^x * q^{n-x})$$

Where p = probability of a success, q = probability of a failure ($1 - p$)

This is really a two-part equation. In the first part, we calculate the binomial coefficient to determine the amount of possible combinations. In the second part, we determine the probability of one of those combinations occurring. Multiplying the two together gives us the total probability of " x " successes in " n " trials. This formula will enable us to calculate probabilities of achieving a specific number of successes given any number of trials.

If we go back to our original problem, the probability of getting exactly 15 heads if we flip a coin 20 times is:

$$15504 * \left(\frac{1}{2}\right)^{15} * \left(\frac{1}{2}\right)^5 = 0.0148$$

=1.48%

Slide 8:

The Binomial Probability Distribution (Cont'd)

Example 2

Let's say that your instructor has given you a surprise, 4-question multiple-choice quiz based on the previous class...that you just happened to miss because your bed was too comfortable! So, you decide to go with your multiple-choice mantra: "When in doubt, pick B". If there are 3

possible answers for each question, and there are 4 questions, what are your chances of:

1. Correctly answering 1 question?
2. Correctly answering all 4 questions?
3. Passing the quiz (2 or more correct answers)?
4. Correctly answering none of the questions?

Show Answer

Answer

Since there are 3 possible answers, that means you have a 1 in 3 chance, or $1/3$ (33.33%) probability of a correct selection. Therefore:

1. $n = 4, x = 1, p = 0.333, q = 0.667. P(X = 1 \text{ correct answer}) = \mathbf{39.50\%}$

Since you are calculating the probability of choosing 1 out of 4 correctly, you use $(4C1 = 4$ "choose" 1) and the chances of a correct answer are 1 out of 3 which implies an incorrect answer is 2 out of 3 so..... $P(x = 1 \text{ correct answer}) = (4C1) (1/3)^1 \times (2/3)^3 = 4(1/3) \times (8/27) = 0.3951$ or 39.5%. There is a 39.5% chance that you will guess one correctly.

2. $n = 4, x = 4, p = 0.333, q = 0.667, P(X = 4 \text{ correct answers}) = \mathbf{1.23\%}$

$P(x = 4 \text{ correct answers}) = (4C4) \times (1/3)^4 \times (2/3)^0 = 1(1/81) \times (1) = 1/81$ or 1.23%. It is extremely unlikely that you will get 100% on this quiz.

3. $n = 4, x = 2, 3, 4, p = 0.333, q = 0.667. P(X = \text{at least 2 correct answers}) = \mathbf{40.7\%}$
($0.296 + 0.099 + 0.012$)

$$P(x = 4 \text{ correct answers}) = (4C4) \times (1/3)^4 \times (2/3)^0 = 1.23\%$$

$$P(x = 3 \text{ correct answers}) = (4C3) \times (1/3)^3 \times (2/3)^1 = 9.9\%$$

$$P(x = 2 \text{ correct answers}) = (4C2) \times (1/3)^2 \times (2/3)^2 = 29.6\%$$

or instead of calculating 3 probabilities, another way of finding the answer is to calculate $P(x = \text{at least 2 correct answers})$

$$= 1 - [P(X = 0) + P(X = 1)]$$

$$= 1 - [(4C0) \times (1/3)^0 \times (2/3)^4 + (4C1) \times (1/3)^1 \times (2/3)^3]$$

$$= 1 - [0.1975 + 0.3951]$$

$$= 1 - 0.5926$$

$$= 0.4074 \text{ or } 40.74\%$$

4. $n = 4, x = 0, p = 0.333, q = 0.667. P(X = 0 \text{ correct answers}) = \mathbf{19.75\%}$

$$P(x = 0 \text{ correct answers}) = (4C0) \times (1/3)^0 \times (2/3)^4 = 0.1975$$

Slide 9:

The Normal Distribution

Overview

Early in the 20th century, a scientist by the name of Galton, who was interested in the shapes of statistical distributions, invented a device called a quincunx. This was a "machine" which allowed

him to study the effects of random variation on the outcome of a number of identical trials. Much like the "Plinko" game on the Price is Right, the object is to drop a disc (or bead) from the top of the device, and after going through a series of obstacles, seeing where the disc will land at the bottom. This may not sound too exciting, but after a number of trials, the discs began to pile up in a distinct pattern. If the process is not manipulated in any way, the end result will likely look like this:

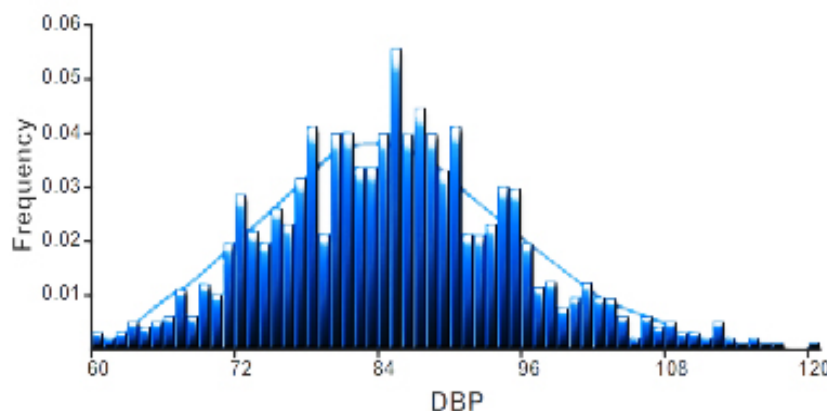


Slide 10:

The Normal Distribution

Overview (Cont'd)

Probability distributions, as we saw in the last section, can take on many different forms. Their shapes are altered by the amount of trials, the number of possible outcomes, as well as the individual probabilities associated with each event. If a large random sample was taken (minimum of 30 members) for a given random variable (let's say one's blood pressure), the distribution will probably look a little like this:



Although there are a wide variety of possible readings, notice that the majority of individuals fall in the middle of the distribution. The further one strays from the "middle" of the graph, the less frequently these individual values will appear.

In both situations, the graph takes on a "bell-curve" appearance, typically characterised by its symmetry and central tendency. This is known as a **normal distribution**.

Slide 11:

The Normal Distribution

Example 1

Let us take a look at another example. Suppose we toss a pair of dice and observe the sum of the two. There are several ways to achieve certain results. For example, to roll a total of 5, we could have: {1 and 4, 4 and 1, 2 and 3, 3 and 2}. In other words, there is a total of 4 situations where the sum would be "5". In fact, the probability distribution for the sum of the dice is as follows:

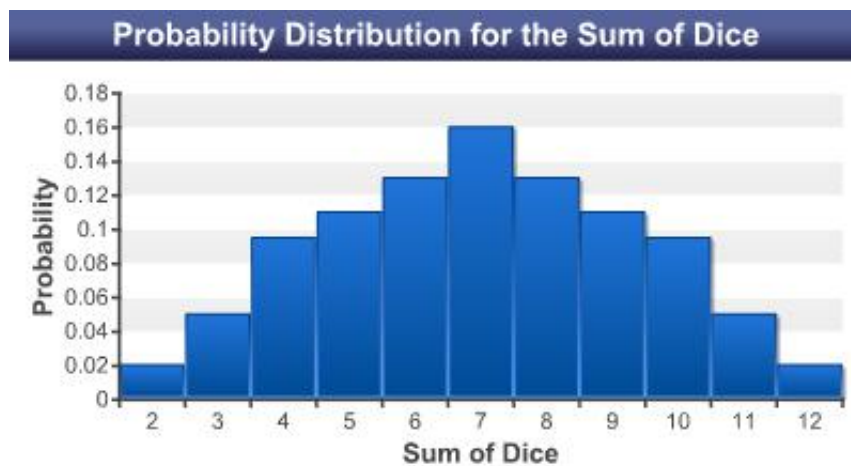
X	P(X)
2	1/36
3	2/36
4	3/36
5	4/36
6	5/36
7	6/36
8	5/36
9	4/36
10	3/36
11	2/36
12	1/36



dice_table

Where **X** is the sum of the dice **p(X)** is the probability of actually obtaining the desired sum. For example, if we want a sum of 5, there are 4 ways of doing it. The probability of rolling a sum of 5 is 4 divided by the total number of possible outcomes (36).

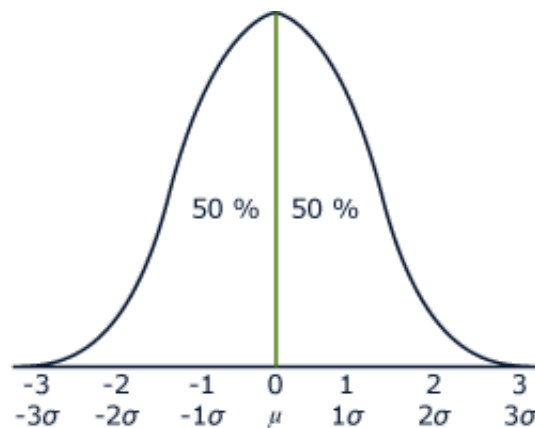
Graphing this probability distribution will yield a familiar pattern:



Slide 12:

Properties of the Normal Distribution

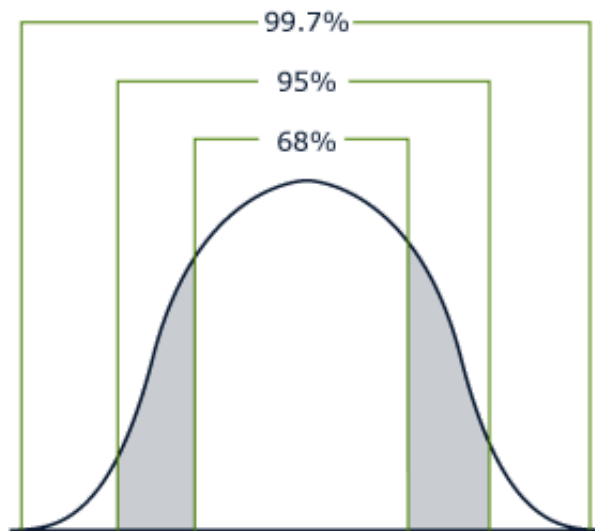
1. The distribution, which is bell-shaped and symmetric, extends infinitely in both directions, approaching, but never touching the horizontal axis.
2. The total area underneath the curve is equal to 1.
3. The mean divides the area in half, leaving 0.50 in each side.
4. The standard normal distribution has a mean of 0 and a standard deviation of 1.
5. Most of the values are found within 3 standard deviations from the mean.



Slide 13:

Empirical Rule

If it is known that a set of values are normally distributed, then the highest point of the curve represents the mean, median, and mode of the data set. Furthermore, within one standard deviation from that mean lie about 68% of the data. 95% of the data will fall within 2 standard deviations, and 99.7% will be contained within 3 standard deviations of the mean. This is known as the Empirical Rule. It applies not only to perfect bell-shaped distributions, but also to any "mound-shaped" graphs that seem to have similar properties.



When referring to the **standard normal distribution**, we are talking about a perfect bell-shaped graph with a mean of 0 and a standard deviation of 1, just like in the graph above. In fact, we can describe the situation using a specialised notation:

$$X \sim N(\mu, \sigma^2)$$

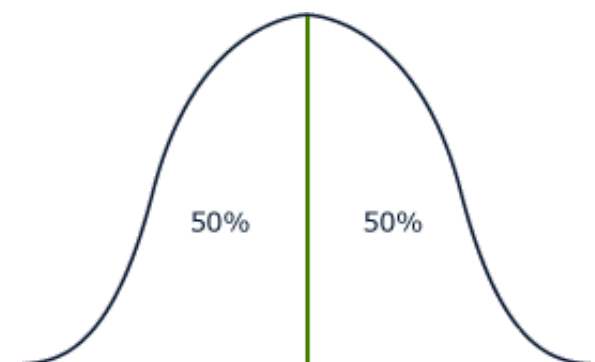
We read this as, "X is normally distributed with a mean of μ and a variance of σ^2 ." In our case, with a standard normal distribution, we would describe the data as: $X \sim N(0, 1)$. The most common mistake that is made here is that students forget that it is the variance that is given within the notation, not the standard deviation, although we can easily calculate one if we have the other. For example, $X \sim N(15, 9)$ represents a data set with a mean of 15 and a standard deviation of 3 (variance of 9).

So what's the big deal about getting a normal distribution? How can that help us in finding probabilities? The secret lies underneath the curve...

Slide 14:

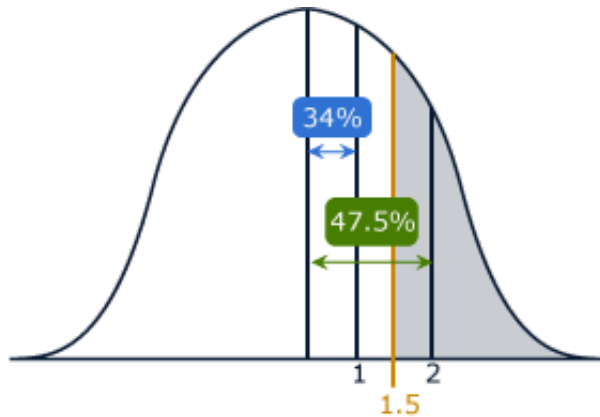
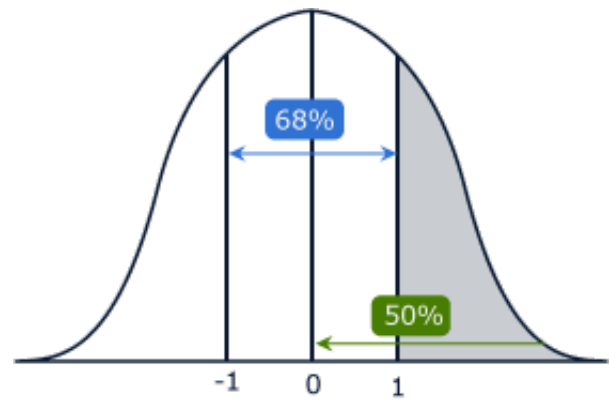
Finding Probabilities using the Normal Distribution

Given a normal distribution with a known mean and variance (standard deviation), we can calculate an endless number of probabilities using the area found under the curve. The rationale is that the curve contains 100% of the area, with half on one side of the mean and half on the other.



Theoretically speaking, we should be able to determine the area found to the right of, to the left of, or between any given data points on a standard normal distribution. For example, if we want the probability of an event being greater than +1 standard deviations away from the mean, $P(X > 1)$, we could easily estimate this using the Empirical Rule.

If we know that 68% of the data falls within 1 standard deviation of the mean, and 50% of the data lies in either side of m , the shaded region should represent $50\% - \frac{1}{2}(68\%) = \mathbf{16\%}$. Therefore, there is a 16% chance that an event will be greater than $+1$. Since the distribution is symmetrical, there is also a 16% chance that the event will be less than -1 .



The Empirical Rule is all we need when dealing with known standard deviation values (0, 1, 2, 3). But what do we do when encountered with a SD of 1.5? Do we take the halfway point between 1 and 2? Since this is a bell-shaped curve, the answer is NO! Using some complicated formulas and a lot of patience, mathematicians were able to calculate the area under the curve with an amazing amount of precision for any standard deviation value between -5.00 and 5.00 (although we won't use them all). They tabulated their results and created our new friend, the [normal distribution table](http://resource_centre/images/Tables/normal.cfm) (select this link to download the table).

Slide 15:

The Normal Distribution Table

Therefore, if you wanted to find the area to the left of 1.92, we would go down the left column until we got to 1.9. Then we would follow that row until it corresponded with 0.02 on the top row. That intersection represents the area found to the left of 1.92. The arrows are there to remind us that the chart is much bigger than the sample you see here.

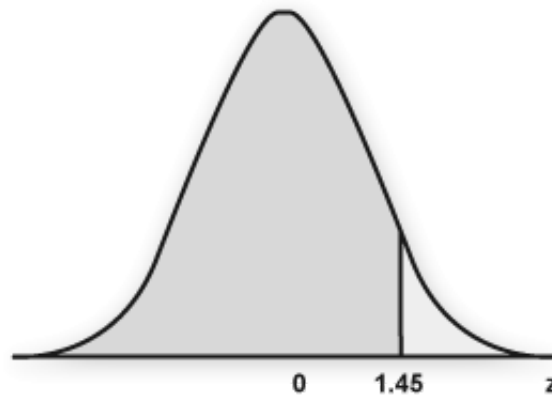
	0.00	0.01	0.02	0.03	0.04
1.0	<p>"z" represents the amount of standard deviations from the mean. The left column is for the first two digits, and the top row represents the last digit of the "z" value. The value found at the intersection of the row and column represents the area to the left of (less than) the z-value.</p>				
1.1					
1.2					
1.3					
1.4					
1.5					
1.6					
1.7					
1.8					
1.9					
2.0					

Let us not forget about our original problem: what is the area to the right of 1.5, or, $P(Z > 1.5)$? According to the piece of the normal distribution table that we have here, the area to the left of

1.5 is 0.9332. Since we want the area to the right of 1.5, simply subtract 0.9332 from 1 to get $(1 - 0.9332) 0.0668$, or 6.7%.

Note: Unlike some other normal distribution tables that you may come across, this table is designed to ALWAYS give the area to the LEFT of the value you look up.

Example: Determine the area under the standard normal curve to the left of $Z = 1.45$ $P(Z < 1.45)$



A portion of the Normal Distribution Table

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06
⋮							
1.4						0.9265	
⋮							

$$P(Z < 1.45) = 0.9265$$

That means that the leftover area $(1 - 0.9265)$ represents the area to the right of 1.45.

$$P(Z > 1.45) = 1 - P(Z < 1.45) = 1 - 0.9265 = 0.0735$$

Slide 16:

The Z-Score

Overview

Easy enough, right? Well, at least it is when dealing with a standard normal distribution with a mean of 0 and a variance of 1. Unfortunately, in practice, this is rarely the case. When faced with data that is normally distributed, but not standardised [not $X \sim N(0, 1)$ = the mean does not equal 0 and the standard deviation is not 1], we must convert the data so that it can be compared to the standard normal distribution. In other words, we must find a method that will convert any given value "x" and express it in terms of standard deviations away from the mean. This is achieved using the standardised **z-score** formula.

$$Z = \frac{x - \mu}{\sigma}$$

Where Z stands for the number of standard deviations a value is from its mean (standard score), X represents that individual value being investigated, μ is the mean of the data set, and the standard deviation is σ .

When you transform your scores into z-scores, the distribution itself does not change. The two things, however, do change:

1) The value of the mean will now equal 0.

2) The value of the standard deviation will now equal 1.

This means that we can now make use of the standard normal distribution table to solve any probability problem relating to this distribution.

Slide 17:

The Z-Score

Example 1

Given a data set with a mean of 12 and a standard deviation of 4, let's find the z-score for the following values:

1. 10
2. 20
3. 16
4. 8

Answer:

1. $Z = (10 - 12) / 4 = -0.5$
2. $Z = (20 - 12) / 4 = 2$
3. $Z = (16 - 12) / 4 = 1$
4. $Z = (8 - 12) / 4 = -1$

Basically, no matter what the mean and variances are, we will always be able to compare our data to the standard normal distribution using the z-score formula as long as a normal distribution exists. This formula will serve as the basis for similar equations that we will encounter in later lessons, especially with hypothesis testing.

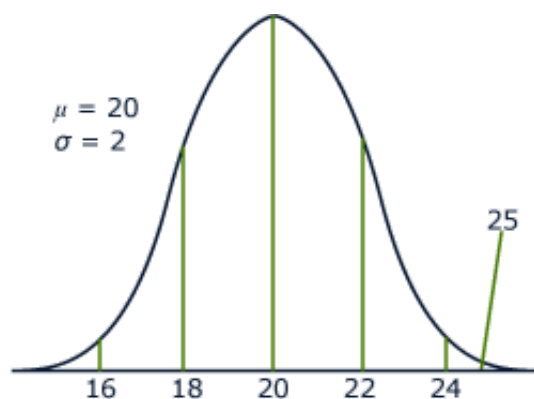
Slide 18:

The Z-Score (Cont'd)

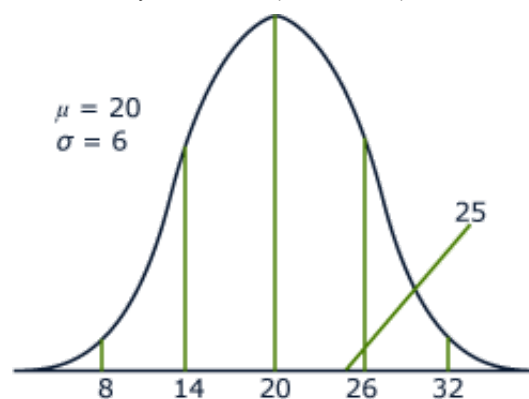
When we have knowledge of both the mean and standard deviation of a population, we can locate where any score falls within a distribution.

Example 2

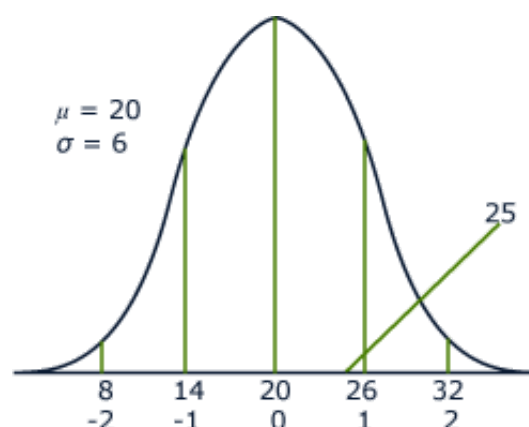
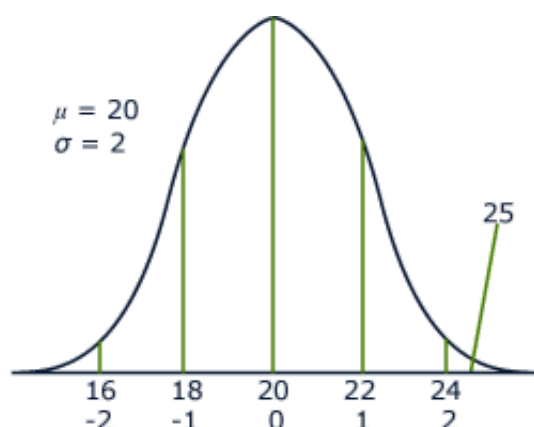
Let's say we are interested in locating where a score of 25 is located in 2 different populations. Both population A and population B are made up of all possible scores from their respective populations.



Population A - In this distribution our score of 25 falls more than 2 standard deviation units away from the mean. Relative to the other scores it falls quite far from the mean and it is positioned in the upper end of the distribution.



Population B - In this distribution our score of 25 falls between the mean and the 1st standard deviation unit. Relative to the other scores it falls quite close to the mean and it is positioned close to the center of the distribution.



Notice that in both cases our mean is now equal to 0 and the standard deviation equals 1. Let's look at the formula for the z-score so that we can see why this is the case.

$$Z = \frac{x - \mu}{\sigma}$$

Recall that when we calculated our standard deviation we first calculated the difference between the random variable (X), and the mean of the data set (μ). So our numerator tells us where the score (X) is located relative to the mean (μ). The denominator (σ) is the value of 1 standard deviation unit in its original form. This formula employs the information needed to end up with a z-score which is based on the original scores deviating from the mean and how that relates to the 'standard' deviation from the mean.

A z-score is a measure of the distance away from the mean in standard deviation units.

Slide 19:

The Z-Score (Cont'd)

So why does μ in a standardized z-score distribution = 0; and why does $\sigma = 1$?

Let's use our population A and population B as examples.

In population A, $\mu = 20$ and $\sigma = 2$. A score of 20 then falls at the mean, a score of 22 falls 1 standard deviation above the mean and a score of 18 falls 1 standard deviation below the mean. If we plug these values into our formula, we get the following:

align="center" cellpadding="0" cellspacing="0" bordercolor="#063763" id="tableData">

X =	18	20	22
(X - μ)	-2	0	2
z-Score	-1	0	1

Using the same example as above, let's locate our score of 25. Recall $X = 25$ fell more than 2 standard deviation units above the mean. Plugging this score into the z-score formula, we get the following:

$$\begin{aligned} Z &= (25 - 20) / 2 \\ &= 5 / 2 \\ &= 2.5 \end{aligned}$$

Here we see that, as in the original raw score form, when $X = 25$ in this particular distribution it falls exactly 2.5 standard deviation units above the mean.

Let's look at population B. We had $\mu = 20$ and $\sigma = 6$. In the original raw scores, 20 is the score at the mean, 26 is one standard deviation above the mean and 14 is one standard deviation below the mean.

align="center" cellpadding="0" cellspacing="0" bordercolor="#063763" id="tableData">

X =	14	20	26
(X - μ)	-6	0	6
z-Score	-1	0	1

Now let's see where the score of 25 falls in this distribution. Remember that in the original raw score form $X = 25$ fell between the mean and the 1st standard deviation unit. As a z-score it will be the following:

$$\begin{aligned} Z &= (25 - 20) / 6 \\ &= 5 / 6 \\ &= 0.83 \end{aligned}$$

We see here that when $X = 25$ in this distribution, it falls exactly 0.83 of a standard deviation unit above the mean.

Using this method of standardising a distribution of scores, anyone who knows what a z-score means can locate where any score falls without knowledge of the actual raw values.

Slide 20:**The Z-Score (Cont'd)****Example 3**

You have just received your grade for a statistics exam, and your professor has deviously given back your score as a z-score. Luckily, he has also given you the mean ($\mu = 70$) and standard deviation of the scores $\sigma = 5$. You got a $Z = 1.3$. Well, you know you did pretty good compared to your classmates since you scored 1.3 standard deviation units above the mean. However, you want to know exactly what your score was, so by rearranging the z-score formula you figure it out.

$Z = (x - \mu) / \sigma$, so $z\sigma = x - \mu$ and $z\sigma + \mu = x$, therefore:

$$\begin{aligned} X &= \mu + Z \sigma \\ X &= 70 + 1.3 (5) \\ X &= 70 + 6.5 \\ X &= 76.5 \text{ (Not Bad!)} \end{aligned}$$

Let's say your score had been $Z = -1.6$.

$$\begin{aligned} X &= 70 - 1.6 (5) \\ X &= 70 - 8 \\ X &= 62 \text{ (Not So Good!)} \end{aligned}$$

Notice that the sign of the z-score tells you if the score falls above or below the mean.

Slide 21:**Exercise 1a**

Given that a normal distribution has a mean of 40 and a variance of 100, please find **$P(X < 35)$** .

Answer:

-Step 1: **Identify the population values.**

Given a variance of 100, the standard deviation is 10. The population mean is 40.

-Step 2: **Convert the random variable (X) into a z-score using the conversion formula.**

$$Z = \frac{x - \mu}{\sigma}$$

$$z = (35 - 40) / 10 = -0.5.$$

-Step 3: **Rewrite the probability in terms of Z:** $P(Z < -0.5)$.

-Step 4: **Find the desired probability using the Normal Distribution Table.**

Using the normal distribution table, we want to find the area to the **left** of -0.5 since we want to know the probability of z being **smaller** than -0.5 .

According to the table, $P(Z < -0.5) = 0.3085$ or 30.85%.



TIP: The "<" tells us that we are looking to the left, just like the head of the arrow pointing that way. Conversely, the ">" would be to the right and would signify a value that is greater than Z).

Slide 22:

Exercise 1b

Given that a normal distribution has a mean of 40 and a variance of 100, please find **P(X > 52)**.

Answer:

-Step 1: **Identify the population values.**

Given a variance of 100, the standard deviation is 10. The population mean is 40.

-Step 2: **Convert the random variable (X) into a z-score using the conversion formula.**

$$Z = \frac{x - \mu}{\sigma}$$

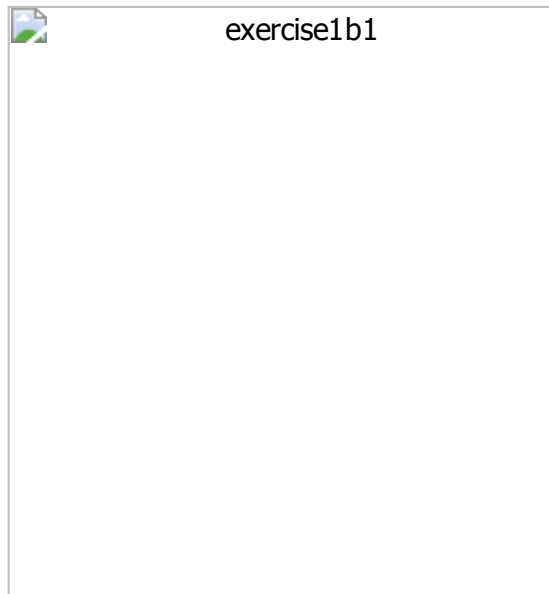
$$z = (52 - 40) / 10 = 1.2.$$

-Step 3: **Rewrite the probability in terms of Z:** P(Z > 1.2).

-Step 4: **Find the desired probability using the Normal Distribution Table.**

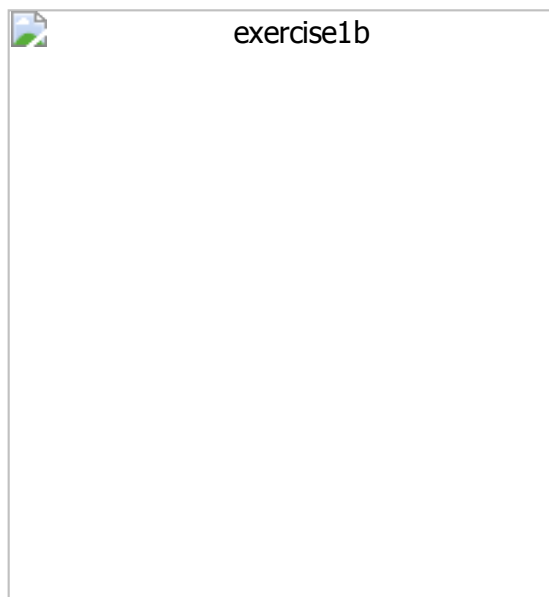
Using the normal distribution table, we want to find the area to the **right** of 1.2 since we want to know the probability of z being **greater** than 1.2 .

According to the Normal Distribution table, when we look up 1.2 we get a value of 0.8849. This does NOT make sense if we consider the relative size of the area that we are trying to identify (see image below).



The shaded area represents the section of the normal distribution that is greater than a z-score of 1.2 and this clearly does not represent 88.49% of the data. Recall that the normal distribution table that we are using always gives us the area to the left of the z-score (less than). If we would like to isolate the area to the right we must subtract 0.8849 from 1 (to get the opposite side of the graph).

In other words, $P(Z > 1.2) = 1 - P(Z < 1.2) = 1 - 0.8849 = 0.1151$ (11.51%).



TIP: Because the normal distribution is symmetric, you could also have found your answer by looking up "-1.2" on the normal distribution table. The area to the right of 1.2 is the same as the area to the left of -1.2/

Slide 23:

Exercise 1c

Given that a normal distribution has a mean of 40 and a variance of 100, please find **$P(20 < X < 66)$** .

Answer:

-Step 1: **Identify the population values.**

Given a variance of 100, the standard deviation is 10. The population mean is 40.

-Step 2: **Convert the random variable (X) into a z-score using the conversion formula.**

$$Z = \frac{x - \mu}{\sigma}$$

z (for X = 66) = (66 - 40) / 10 = 2.6.

z (for X = 20) = (20 - 40) / 10 = -2.0.

-Step 3: **Rewrite the probability in terms of Z:** $P(-2.0 < Z < 2.6)$.

-Step 4: **Find the desired probability using the Normal Distribution Table.**

There are several techniques that can be used to determine an "in-between" probability. However, considering the nature of the normal distribution table used in this course, the most efficient way of determining the probability is by starting with the area for the largest value (2.6) and subtracting from it the area for the smaller one (-2.0):

- $P(Z < 2.6) - P(Z < -2.0)$
- Area to the left of 2.6 = 0.9943
- Area to the left of -2.0 = 0.0228
- $0.9943 - 0.0228 = 0.9725$ (97.25%).



exercise1c

Slide 24:

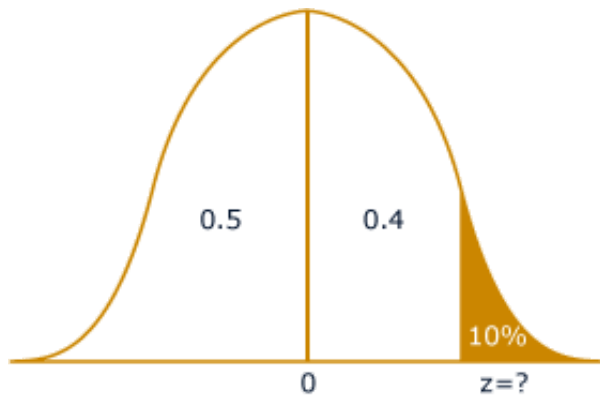
Exercises 2

Your instructor tells you that you need to be in the top 10% of the class to get an A on the final exam. From the previous exams that he's given, you've found that a normal distribution exists with a mean of 72, and a standard deviation of 13. What is the minimum grade that you'll need on your final exam to get that A?

Answer:

This question is a little different since we now want to find "X" given "Z" instead of the other way around. We're not exactly given Z on a silver platter, but since we know the probability (10%), we

can work our way backwards to get it.



We want to find the z-value where 90% of the data falls to the left of it (10% falls to the right).

This means that we have to find 0.9000 on the [normal distribution table](http://resource_centre/images/Tables/normal.cfm). Then we work our way back to identify the z-value.

From the Normal Distribution Table

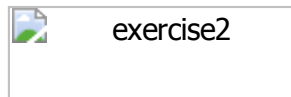
1.0	0.8413	0.8438	0.8465	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9236	0.9251	0.9265	0.9278	0.9292	0.9306	0.9319

closest to 0.9000

This table represents a portion of the Normal Distribution Table. To download the entire table, you may use the drop-down menu found above ("Select Table - Normal Distribution").

The closest thing we have to 0.9000 is the 0.8997. If we work our way back, we find that it corresponds to a z-value of 1.28. Using the z-score formula, we can substitute the "z" for 1.28, the μ for 72, and the σ for 13.

$$Z = \frac{x - \mu}{\sigma}$$



Solving for x:

$$1.28(13) = x - 72$$

$$16.64 = x - 72$$

$$x = 16.64 + 72$$

$$x = 88.64\%$$

Therefore, we will need at least 89% on the final exam to get that well-earned "A".

Slide 25:

Exercise 3

According to the normal distribution table, how accurate is the empirical rule with respect to the amount of data contained within 1, 2, and 3 standard deviations from the mean?

Answer:

It is very accurate, since according to the normal distribution, there is 68.26% of the data within 1 standard deviation (empirical rule = 68%), 95.44% within 2 SDs (95% with empirical), and 99.74% within 3 SDs (99.7% with E.R.).

Slide 26:

Uses of Z-Scores

One of the many uses for z-score transformation is the ability to compare two scores from two different distributions. For example, we have a subject who participated in an experiment that was looking at abilities in attention. This subject obtained a score of $X = 20$ on one measure of attention and a score of $X = 42$ on a second measure of attention. Because these are two different measures, one cannot directly compare the two scores. Indeed, one score could be at the top of the distribution and one could be on the bottom end. We can compare these two scores if we first transform them into z-scores.

Measure # 1

$$X = 20, \mu = 10, s = 4$$

Measure # 2

$$X = 42, \mu = 20, s = 10$$

For measure #1

$$Z = (20 - 10) / 4$$

$$Z = 2.5$$

For measure #2

$$Z = (42 - 20) / 10$$

$$Z = 2.2$$

From this information we can tell that this subject scored quite high on both of these measures as compared to the other participants. We know this because in both cases this subject's scores are over 2 standard deviation units above the mean in each respective distribution.

Another important use of the z-score is the ability to figure out (in most cases) the probability of obtaining a specific score given the makeup of that specific distribution.



Video Examples

For the past 20 years, the Medical College Admissions Tests (MCAT) has been given to students who have aspired for an MD. Upon review of the statistics of this exam, you have noticed that the individual scores are normally distributed with a mean of 7.87, and a standard deviation of 1.83. What is the probability of scoring:

- Less than 11?
- Over 9?
- Between 5.5 and 8.5?

Solutions

The video explains how to use the normal distribution to calculate probabilities. (Length 17:32)

- View [streaming video](#) if you have a high-speed Internet connection.

- [Download the video](#) (101 Mb) if you have a low-speed Internet connection.



Click for Bonus

Slide 27:

Recap

Several tools and procedures exist in statistics that help us determine probabilities if said data are presented in a given way.

- The binomial distribution function (formula) helps us calculate probabilities of binomial events, and this is especially useful when the number of trials is such that finding each possibility one by one is too time-consuming.
- The probability of any random variable can be determined if one knows the mean and variance of the population from whence it came, and that population is normally distributed.
- The z-score formula converts any given variable into a standardised score, given that one knows the mean and variance (standard deviation) of the population under study.

You can post a message online in your discussion folder any time you have something to share with your discussion group concerning the current lesson. Simply click [Discussion Board](#) or use the menu at the top of the screen.

Next lesson: Estimates and Repeated Sampling